SICACT NEWS

PUBLISHED BY THE ACM SPECIAL INTEREST COMMITTEE
ON AUTOMATA THEORY AND COMPUTABILITY

No. 4 JANUARY 1970

CHAIRMAN:

Patrick C. Fischer
Department of Applied Analysis
and Computer Science
Waterloo, Ontario, Canada

VICE-CHAIRMAN:

Michael A. Harrison
Department of Computer Science
University of California
Berkeley, California 94720

SECRETARY:

A. V. Aho
Bell Telephone Laboratories
Murray Hill, New Jersey 07971

CONTENTS

Chairman's Column 1
Meetings Ahead. 2
Research Notes. 3
Recent Publications 19

FROM THE NEWSLETTER EDITOR:

With this issue, SICACT News has become self-supporting in the sense that editing merely meant assembling what had been sent to me. I would like to thank all the contributors and encourage others to participate in making SICACT News an active publication.

Jurg Nievergelt
Department of Computer Science
University of Illinois at
Urbana-Champaign
Urbana, Illinois 61801

CHAIRMAN'S COLUMN

According to Mrs. Irene Hollister of ACM-HQ, we have received the necessary number of petitions to ratify our conversion to a SIG. Official approval now must be given by the ACM Council. In order not to delay the conversion a mail ballot is being circulated to the Council.

The annual dues will initially be $2.50 for ACM members and $5.00 for ACM non-members. Non-member dues will be payable beginning April 1, while ACM members will not be asked to pay dues until their next renewal of ACM membership.

As is inevitable in a mail solicitation, quite a few members of SICACT did not respond to the conversion mailing. This could have been due to the Christmas mail rush, lack of interest in SIGACT or other reasons. If you did not respond to the December conversion mailing, you must now reaffirm your desire to be in SIGACT in order to remain on the membership rolls. This can be done now simply by filling out and sending in the form at the bottom of this page. Persons from whom we cannot obtain some sort of a response by the end of March will be dropped from the Group. If you are dropped by this procedure, you are of course welcome to rejoin at any time, but to do so will require at least as much effort as detaching and mailing the form below right now. Thanks very much for your cooperation.

Patrick C. Fischer.

ACM Special Interest Group on Automata and Computability Theory

Mail to: Dr. A. V. Aho, Secretary of SIGACT
Bell Telephone Laboratories
Murray Hill, N.J. 07971

Date: _____

I wish to reaffirm my interest in SIGACT and to be a member.

Name _____ Signature: _____
(please print)

( ) I am a member of ACM. My membership number is _____.

( ) I am not a member of ACM. My address is:

_____

_____

_____

SECOND ANNUAL SYMPOSIUM ON THEORY OF COMPUTING

Out of more than 65 submissions, 30 papers have been selected for the 1970 Second Annual ACM Symposium on Theory of Computing. The complete program will be published in the next issue of the SIGACT News. A brief summary of the Symposium is given below.

Location: Colonial Hilton, Northampton, Mass. 01060.
Supporting Institution: Smith College.
Local Arrangements: Prof. Robert Fabian, Clark Science Center, Smith College.
Publicity: Dr. Jeff Ullman, Dept. of Electrical Eng'g, Princeton University.

Sunday, May 3: Registration, Hospitality Room
Monday, May 4: Session I (Abstract Complexity Theory)
Session II (Parsing, Optimization of Programs)
Business Meeting of SIGACT
Tuesday, May 5: Session III(AFL's, Tree Structures)
Session IV (Logic, Theorem Proving)
Cocktail Reception
Wed., May 6: Session V (Algebraic Structures)

A Note on Independence of a Regularity-Preserving Operator

by

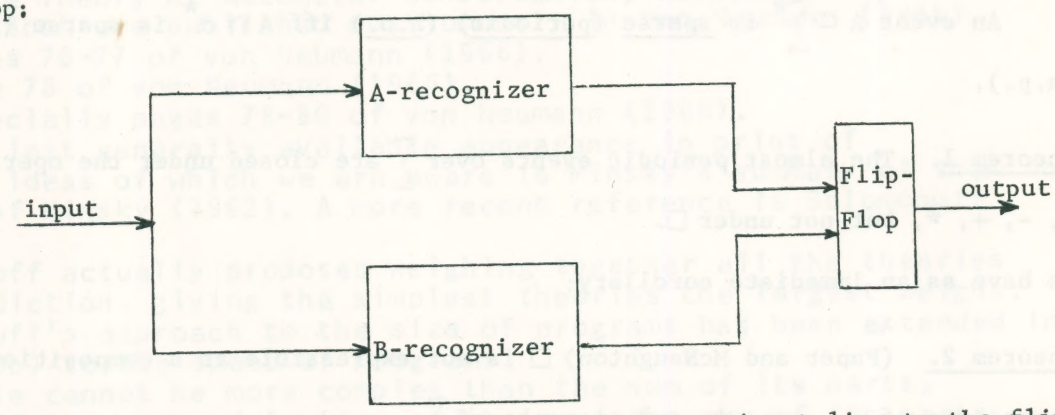
M. Fischer, A. Meyer, P. O'Neil, and M. Paterson

The "box" operator on languages (sets of strings) is introduced by Papert and McNaughton [1.] in their monograph on star-free regular events. If $\sum$ is a finite alphabet, $A,B \subset \sum^*$ = the free semigroup generated by $\sum$, then $A \square B$ is the set of words in $\sum^*$ which have some prefix in A, and whose longest prefix in A is strictly longer than any prefix in B, viz.,

Definition. For $A,B \subset \sum^*$,

$$A \square B = \{xy \mid x \in A,\ y \in \sum^*, \text{ and } (\forall u,v)[y = uv \Rightarrow xu \notin B]\}$$

The box operator is most naturally described as the language accepted by a parallel connection of an A-recognizer and B-recognizer in series with a flip-flop:



An input word is in $A \square B$ if and only if the upper input line to the flip-flop received a signal after the last signal on the lower line. It is obvious from this description that if A and B are regular events, then so is $A \square B$. Moreover, appealing to the theorem ([1.], [2.]) that an event is star-free if and only if it is accepted by a cascade connection of flip-flops, it becomes obvious that if A and B are star-free regular events, then so is $A \square B$.

This latter observation will not follow by an elementary argument from the definition of star-free events, because box cannot be expressed in terms of the star-free operators "·" (concatenation), "+" (union), and "-" (relative complementation). McNaughton and Papert in an unpublished paper have proved the stronger result that even with the use of "*" (Kleene closure) box cannot be expressed.

In this note we prove an extension of their theorem by exhibiting a family of languages closed under ·, +, -, and *, but not closed under □. Our proof uses ideas introduced by Papert and McNaughton, but is somewhat simpler.

Definition. Let $\sigma \in \sum$ be fixed. An event $A \subset \sigma^*$ is sparse iff for some polynomial f, $|\{x \in A \mid \text{length}(x) \leq n\}| \leq f(\log_2 n)$ for all integers n. An event $A \subset \sigma^*$ is periodic iff for some integer p, $\{\text{length}(x) \mid x \in A\}$ is a union (possibly empty) of congruence classes of nonnegative integers modulo p. An event $A \subset \sigma^*$ is almost periodic (a.p.) if for some periodic set $A' \subset \sigma^*$, and some sparse set S, $A'-S \subset A \subset A'+S$.

An event $A \subset \sum^*$ is sparse (periodic) (a.p.) iff $A \cap \sigma^*$ is sparse (periodic) (a.p.).

Theorem 1. The almost periodic events over $\sum$ are closed under the operations of ·, -, +, *, but not under □.

We have as an immediate corollary:

Theorem 2. (Paper and McNaughton) □ is not expressible as a composition of the operations of ·, -, +, *, and elements of $\sum$.

Theorem 1 follows from the following observations:

Fact 1. $(\sigma^* \cap A) \circ (\sigma^* \cap B) = \sigma^* \cap (A \circ B)$ when $\circ$ is replaced by ·, -, or +. Also, $(\sigma^* \cap A)^* = \sigma^* \cap A^*$. We conclude that the closure of the a.p. events contained in σ^* under ·, -, +, and *, implies the closure under these operations of the a.p. events contained in $\sum^*$. Thus we henceforth assume all events are contained in σ^*.

Fact 2. If A is an a.p. event, then $\lim_{n\to\infty}(|\{x \in A \mid \text{length}(x) \leq n\}|/n)$ exists. This is obvious if A is periodic, and if A differs from a periodic set by a sparse set the limit is unaffected.

Fact 3. Let $A = \{\sigma^k \mid k = 2^n \text{ for some } n \geq 1\}$ and $B = \{\sigma^k \mid k = 3\cdot 2^n \text{ for some } n \geq 0\}$. Then $A \square B = \{\sigma^k \mid 2^{n+1} \leq k < 2^n + 2^{n+1} \text{ for some } n \geq 0\}$ and $\lim_{n\to\infty}(|\{x \in A \square B \mid \text{length}(x) \leq n\}|/n)$ does not exist. A and B are clearly sparse, and sparse events are a.p. Hence, fact 2 implies the a.p. events are not closed under $\square$.

Fact 4. Periodic events are obviously closed under complementation relative to σ^*. Since $A'-S \subset A \subset A'+S$ implies $\overline{A}'+S \supset \overline{A} \supset \overline{A}'-S$, it follows that A is a.p. iff $\overline{A}$ is a.p. Hence, a.p. events are closed under complementation relative to σ^*.

Fact 5. For any event A, the event A^* is ultimately periodic. Namely, $A^* = B-F$ where B is a periodic event with period equal to g.c.d. $(\{\text{length}(x) \mid x \in A\})$ and F is finite.‡ Since finite sets are sparse, we conclude that the a.p. events are closed under *.

Fact 6. Sparse sets are closed under + (the polynomials add), as are periodic sets (with period equal to the l.c.m. of the periods). It follows easily that a.p. events are closed under +.

Fact 7. Let A be an a.p. event and k be any multiple of the period. (If A is sparse, choose $k > 0$). Let A_r be $A \cap \sigma^r(\sigma^k)^*$, the words in A of length congruent to r modulo k. Then $A_r = \sigma^r C_r$ for some $C_r \subset (\sigma^k)^*$ such that either C_r is sparse, or $(\sigma^k)^* - C_r$ is sparse.

---

‡An event A is ultimately periodic iff $A'-F \subset A \subset A'+F$ for some periodic event A' and finite set F. It may be helpful to note that fact 5 follows from the more easily proved observation that the period of A^* divides the length of the shortest word in A of positive length.

Let A, B be a.p. events and k be the l.c.m. of their periods. So $A = \bigcup_{0\le r<k} A_r$, $B = \bigcup_{0\le r<k} B_r$, and $A\cdot B = \bigcup A_i\cdot B_j$. From facts 6 and 7, we conclude that $A\cdot B$ will be an a.p. event if we verify:

Fact 8. Let $C,D \subset (\sigma^k)^*$ be events which are sparse or whose complement relative to $(\sigma^k)^*$ is sparse; then $C\cdot D$ is an a.p. event.

To prove this, note that the sparse events are closed under $\cdot$ (the polynomials multiply). Thus, we assume without loss of generality that $S = (\sigma^k)^* - C$ is sparse. Choose $\sigma^{kn} \in D$ (if $D = \phi$, the result is immediate). Then

$$(\sigma^k)^* \supset C\cdot D \supset C\cdot\sigma^{kn} = ((\sigma^k)^* - S)\cdot\sigma^{kn} = (\sigma^k)^* - (\sigma^{kn}S+F)$$

where F is the finite set of words $\lambda, \sigma^k, \cdots, \sigma^{k(n-1)}$. But $\sigma^{kn}S+F$ is sparse, and of course $(\sigma^k)^*$ is periodic, so $C\cdot D$ is a.p.

This completes the proof.

Remark. Theorem 1 is stronger than Theorem 2. In general it is possible to have operations 0_1, 0_2, 0_3 on a set such that <u>any</u> subset closed under 0_1 and 0_2 is closed under 0_3, but 0_3 is not expressible as a composition of 0_1 and 0_2. (For example, let 0_1, 0_2 be constant functions and 0_3 non-constant such that range range $(0_3) \subset$ range $(0_1) \cup$ range (0_2).)

Remark. If A,B are regular, then so is $A \square B$, and hence $A \square B$ can be expressed in terms of +, $\cdot$, * and letters of $\sum$ alone, though the expression of course depends on A and B. This suggests the question whether there is an expression E involving $\cdot$, -, +, *, letters of $\sum$ and variables X,Y such that the identity $E(X,Y) = X \square Y$ is valid for all regular events X,Y, or even for all finite sets X,Y. McNaughton and Papert show that any such identity valid for all finite sets must be valid for all sets, and hence by Theorem 1 no such identity can exist.

Remark. Theorem 1 can be strengthened to show the independence of $\square$ from additional operations as well. There are clearly a large number of such independence results yet to be catalogued. One question we have been unable to settle is whether $\square$ is independent of $\cdot$, $+$, $-$, $*$ together with left and right quotients of languages.

References

1. Papert, S. and McNaughton, R.: Counter-Free Automata, to appear (1970).

2. Meyer, A. R.: A Note on Star-free Events. J.A.C.M., 16, 2 (April, 1969), pp. 220-225.

Massachusetts Institute of Technology

November, 1969

Events which are not representable by a probabilistic automaton

by

A. Paz

Technion IIT and U.C.L.A.

Let $A = (S,\pi,\{A(\sigma)\},\eta)$ be a probabilistic automaton, S is the (finite) state set, π and η are the initial and final vectors respectively over the states and $\{A(\sigma)\}$ are the transition matrices corresponding to the letters $\sigma \in \Sigma$. If $x = \sigma_1 \ldots \sigma_k \in \Sigma^*$ then $A(x)$ is the matrix $A(x) = A(\sigma_1) \ldots A(\sigma_k)$ and $p^A(x)$ is the word-function with domain Σ^* and real values in the interval $[0,1]$: $p^A(x) = \pi A(x) \eta$. Let λ be a real number $0 \leq \lambda < 1$, the notation $T(A,\lambda)$ or $T(p^A,\lambda)$ stands for the language (set of words or event) $T(A,\lambda) = \{x; x \in \Sigma^*, p^A(x) > \lambda\}$. Such a language will be called probabilistic.

Problem: Is there any language L which is not probabilistic?

Bukharaev (1965) has given an example of such a language over a single letter alphabet. Unfortunately, he did not prove his example (which is quite complicated, being defined on the basis of four arbitrary primitive recursive functions). We have not been able to prove Bukharaev's result, nor have we succeeded in establishing contact with him (directly or indirectly) and, to the best of our knowledge nobody else in the West knows how to prove that result (see e.g. Salomaa (1968) who needed that result in order to establish some relation between time variant events and probabilistic events). Note that, when looking for a language as above, one must take into consideration that there may be different automata defining

the same language; in addition, if a single letter alphabet is assumed then the language must be at least context sensitive, for context free languages over a single letter alphabet are regular and therefore are also probabilistic (see Paz (1966)). Here is an example:

Let $\Sigma = \{\sigma\}$ and let $\Delta = \{a,b\}$. Let $u_1,u_2,\ldots$ be a lexicographical enumeration of all nonempty words over Δ, and let X be the infinite sequence of letters from Δ resulting from the concatenation of the words $u_1,u_2,\ldots$ in their given order. Let $X(i)$ denote the i-th letter in X and define the language over Σ $L_1 = \{\sigma^n;\ x(n) = a\}$, then L_1 is not probabilistic. To prove this, assume the contrary, i.e. let $L_1 = T(A,\lambda)$ where $A = (S,\pi,A(\sigma),n)$. It follows from a theorem of Matsura (see Nasu and Honda (1968)) that for any integer K we have that

$$p^A(\sigma^{k+n}) = c_0 p^A(\sigma^k) + c_1 p^A(\sigma^{k+1}) + \ldots + c_{n-1} p^A(\sigma^{k+n-1}) \qquad (*)$$

where $c_0,c_1\ldots c_{n-1}$ are fixed constants, n being the number of states in S.

Let $\varepsilon_0 \ldots \varepsilon_n$, $\delta_0 \ldots \delta_n$ be two words in Δ^* defined as follows: $\varepsilon_i = b$ if and only if $c_i > 0$, $\varepsilon_i = a$ otherwise, $\varepsilon_n = a$, $\delta_i = \Delta - \varepsilon_i$. By the construction of the sequence X, there are integers k_1 and k_2 such that:

$$X(k_1) \ldots X(k_1 + n) = \varepsilon_0 \ldots \varepsilon_n$$
$$X(k_2) \ldots X(k_2 + n) = \delta_0 \ldots \delta_n$$

It follows from the above construction and from the definition of L_1 that:

$$c_i > 0 \Rightarrow X(k_1+i) = b,\ X(k_2 + i) = a \Rightarrow$$
$$\Rightarrow p^A(\sigma^{k_1+i}) > \lambda,\ p^A(\sigma^{k_2+i}) \leq \lambda.$$

$$c_i \leq 0 \Rightarrow X(k_1+i) = a,\ X(k_2+i) = b \Rightarrow$$

$$p^A(\sigma^{k_1+i}) > \lambda,\ p^A(\sigma^{k_2+i}) \leq \lambda,$$

$$X(k_1+n) = a\ ,\ X(k_2+n) = b \Rightarrow p^A(\sigma^{k_1+n}) > \lambda\ ,$$

$$p^A(\sigma^{k_2+n}) \leq \lambda$$

Evaluating now the formula (*) first for $k = k_1$ and then for $k = k_2$ we have

$$\lambda < p^A(\sigma^{k_1+n}) = \sum_{i=1}^{n-1} c_i p^A(\sigma^{k_1+i}) \leq \lambda \sum_{i=1}^{n-1} c_i$$

and

$$\lambda \geq p^A(\sigma^{k_2+n}) = \sum_{i=1}^{n-1} c_i p^A(\sigma^{k_2+n}) \geq \lambda \sum_{i=1}^{n-1} c_i$$

a contradiction! The claim is thus proved.

Remark: It is easy to see that the language L_1 is context sensitive, for the sequence X can be generated by a linear bounded automaton, and, as remarked at the beginning, no single letter and simpler language can satisfy the required properties. On the other hand, by constructing the sequence X in a more complex way (but still containing as finite subsequences all words in Δ^*) one can construct languages which are not probabilistic but not context sensitive either.

Note: The report herein is a preliminary draft and the author intends to explore various ramifications before publishing.

Bibliography

Bukharaev, R. (1965), Criteria for the representation of events in finite probabilistic automata. Dokl. Akad. Nauk SSSR 164, 289-291.

Paz, A. (1966), Some aspects of probabilistic automata. Information and Control 9, No. 1, 26-60.

Nasu, M., and Honda, N., (1968), Fuzzy Events Realized by Finite Probabilistic Automata, Information and Control 12, 284-303.

Salomaa, A. (1968), On Finite Automata with a Time-Variant Structure. Information and Control 13, 85-98.

TO A MATHEMATICAL DEFINITION OF 'LIFE'

G. J. Chaitin

'Life' and its 'evolution' are fundamental concepts that have not yet been formulated in precise mathematical terms, although some efforts in this direction have been made. We suggest a possible point of departure for a mathematical definition of 'life'. This definition is based on the computer and is closely related to recent analyses of 'inductive inference' and 'randomness'. A living being is a unity; it is simpler to view a living organism as a whole than as the sum of its parts. If we want to compute a complete description of a region of space-time that is a living being, the program will be smaller in size if the calculation is done all together, than if it is done by independently calculating descriptions of parts of the region and then putting them together.

Section 1. The problem.--

'Life' and its 'evolution' from the lifeless are fundamental concepts of science. According to Darwin and his followers, we can expect living organisms to evolve under very general conditions. Yet this theory has never been formulated in precise mathematical terms. Supposing Darwin is right, it should be possible to formulate a general definition of 'life' and to prove that under certain conditions we can expect it to 'evolve'. If mathematics can be made out of Darwin, then we will have added something basic to mathematics; while if it cannot, then Darwin must be wrong, and life remains a miracle which has not been explained by science.

The point is that the view that life has spontaneously evolved, and the very concept of life itself, are very general concepts, which it should be possible to study without getting involved in, for example, the details of quantum chemistry. We can idealize the laws of physics and simplify them and make them complete, and then study the resulting universe. It is necessary to do two things in order to study the evolution of life within our model universe. First of all, we must define 'life', we must characterize a living organism in a precise fashion. At the same time it should become clear what the complexity of an organism is, and how to distinguish primitive forms of life from advanced forms. Then we must study our universe in the light of the definition. Will an evolutionary process occur? What is the expected time for a certain level of complexity to be reached? Or can we show that life will probably not evolve?

Section 2. Previous work.--

Von Neumann devoted much attention to the analysis of fundamental biological questions from a mathematical point of view.(1) He considered a universe consisting of an infinite plane divided into squares. Time is quantized, and at any moment each square is in one of 29 states. The state of a square at any time depends only on its previous state and the previous states of its four neighboring squares. The universe is homogeneous; the state transitions of all squares are governed by the same law. It is a deterministic universe. Von Neumann showed that a self-reproducing general-purpose computer can exist in his model universe.

A large amount of work on these questions has been done since von Neumann's initial investigations, and a complete bibliography would be quite lengthy. We may mention Moore (1962), Arbib (1966,1967), and Codd

(1968).

The point of departure of all this work has been the identification of 'life' with 'self-reproduction', and this identification has both helped and hindered. It has helped, because it has not allowed fundamental conceptual difficulties to tie up work, but has instead permitted much that is very interesting to be accomplished. But it has hindered because, in the end, these fundamental difficulties must be faced. At present the problem has evidenced itself as a question of 'good taste'. As von Neumann remarks,(2) good taste is required in building one's universe. If its elementary parts are assumed to be very powerful, self-reproduction is immediate. Arbib (1966) is an intermediate case.

What is the relation between self-reproduction and life? A man may be sterile, but no one would doubt he is alive. Children are not identical to their parents. Self-reproduction is not exact; if it were, evolution would be impossible. What's more, a crystal reproduces itself, yet we would not consider it to have much life. As von Neumann comments,(3) the matter is the other way around. We can deduce self-reproduction as a property which must be possessed by many living beings, if we ask ourselves what kinds of living beings are likely to be around. Obviously, a species that did not reproduce would die out. Thus, if we ask what kinds of living organisms are likely to evolve, we can draw conclusions concerning self-reproduction.

Section 3. Simplicity and complexity.--

'Complexity' is a concept whose importance and vagueness von Neumann emphasized many times in his lectures.(4) Due to the work of Solomonoff, Kolmogorov, Chaitin, Martin-Lof, Willis, and Loveland, we now understand this concept a great deal better than it was understood while von Neumann worked. Obviously, to understand the evolution of the complexity of living beings from primitive, simple life to today's very complex organisms, we need to make precise a measure of complexity. But it also seems that perhaps a precise concept of complexity will enable us to define 'living organism' in an exact and general fashion. Before suggesting the manner in which this may perhaps be done, we shall review the recent developments which have converted 'simplicity' and 'complexity' into precise concepts.

We start by summarizing Solomonoff's work.(5) Solomonoff proposes the following model of the predicament of the scientist. A scientist is continually observing increasingly larger initial segments of an infinite sequence of 0's and 1's. This is his experimental data. He tries to find computer programs which compute infinite binary sequences which begin with the observed sequence. These are his theories. In order to predict his future observations, he could use any of the theories. But there will always be one theory that predicts that all succeeding observations will be 1's, as well as others that take more account of the previous observations. Which of the infinitely many theories should he use to make the prediction? According to Solomonoff, the principle that the simplest theory is the best should guide him.(6) What is the simplicity of a theory in the present context? It is the size of the computer program. Larger computer programs embody more complex theories, and smaller programs embody simpler theories.

Willis has further studied the above proposal, and also has introduced the idea of a hierarchy of finite approximations to it. To my knowledge, however, the success which predictions made on this basis will have has not been made completely clear.

We must discuss a more technical aspect of Solomonoff's work. He

realized that the simplicity of theories, and thus also the predictions, will depend on the computer which one is using. Let us consider only computers whose programs are finite binary sequences, and measure the size of a binary sequence by its length. Let us denote by C(T) the complexity of a theory T. By definition, C(T) is the size of the smallest program which makes our computer compute T. Solomonoff showed that there are 'optimal' binary computers C that have the property that for any other binary computer C', $C(T) < C'(T)+d$, for all T. Here d is a constant that depends only on C and C', not on T. Thus, these are the most efficient binary computers, for their programs are shortest. Any two of these optimal binary computers C1 and C2 result in almost the same complexity measure, for from $C1(T) < C2(T)+d12$ and $C2(T) < C1(T)+d21$, it follows that the difference between C1(T) and C2(T) is bounded. The optimal binary computers are transparent theoretically, they are enormously convenient from the technical point of view. What's more, their optimality makes them a very natural choice.(7) Kolmogorov and Chaitin later independently hit upon the same kind of computer in their search for a suitable computer upon which to base a definition of 'randomness'.

However, the naturalness and technical convenience of the Solomonoff approach should not blind us to the fact that it is by no means the only possible one. Chaitin first based his definition of randomness on Turing machines, taking as the complexity measure the number of states in the machine, and he later used bounded-transfer Turing machines. Although these computers are quite different, they lead to similar definitions of randomness. Later it became clear that using the usual 3-tape-symbol Turing machine and taking its size to be the number of states leads to a complexity measure CT(T) which is asymptotically just a Solomonoff measure CS(T) with its scale changed: CS(T) is asymptotic to 2 CT(T) log2 CT(T). It appears that people interested in computers may still study other complexity measures, but to apply these concepts of simplicity/complexity it is at present most convenient to use Solomonoff measures.

We now turn to Kolmogorov's and Chaitin's proposed definition of randomness or patternless. Let us consider once more the scientist confronted by experimental data, a long binary sequence. This time he is not interested in predicting future observations, but only in determining if there is a pattern in his observations, if there is a simple theory that explains them. If he found a way of compressing his observations into a short computer program which makes the computer calculate them, he would say that the sequence follows a law, that it has pattern. But if there is no short program, then the sequence has no pattern--it is random. That is to say, the complexity C(S) of a finite binary sequence S is the size of the smallest program which makes the computer calculate it. Those binary sequences S of a given length n for which C(S) is greatest are the most complex binary sequences of length n, the random or patternless ones. This is a general formulation of the definition. If we use one of Solomonoff's optimal binary computers, this definition becomes even clearer. Most binary sequences of any given length n require programs of about length n. These are the patternless or random sequences. Those binary sequences which can be compressed into programs appreciably shorter than themselves are the sequences which have pattern. Chaitin and Martin-Lof have studied the statistical properties of these sequences, and Loveland has compared several variants of the definition.

This completes our summary of the new rigorous meaning which has been given to simplicity/complexity--the complexity of something is the

size of the smallest program which computes it or a complete description of it. Simpler things require smaller programs. We have emphasized here the relation between these concepts and the philosophy of the scientific method. In the theory of computing the word 'complexity' is usually applied to the speed of programs or the amount of auxiliary storage they need for scratch-work. These are completely different meanings of complexity. When one speaks of a simple scientific theory, one refers to the fact that few arbitrary choices have been made in specifying the theoretical structure, not to the rapidity with which predictions can be made.

Section 4. What is life?--

Let us once again consider a scientist in a hypothetical situation. He wishes to understand a universe very different from his own which he has been observing. As he observes it, he comes eventually to distinguish certain objects. These are highly interdependent regions of the universe he is observing, so much so, that he comes to view them as wholes. Unlike a gas, which consists of independent particles that do not interact, these regions of the universe are unities, and for this reason he has distinguished them as single entities.

We believe that the most fundamental property of living organisms is the enormous interdependence between their components. A living being is a unity; it is much simpler to view it as a whole than as the sum of parts. That is to say, if we want to compute a complete description of a region of space-time that is a living being, the program will be smaller in size if the calculation is done all together, than if it is done by independently calculating descriptions of parts of the region and then putting them together. What is the complexity of a living being, how can we distinguish primitive life from complex forms? The interdependence in a primitive unicellular organism is far less than that in a human being.

A living being is indeed a unity. All the atoms in it cooperate and work together. If Mr. Smith is afraid of missing the train to his office, all his incredibly many molecules, all his organs, all his cells, will be cooperating so that he finishes breakfast quickly and runs to the train station. If you cut the leg of an animal, all of it will cooperate to escape from you, or to attack you and scare you away, in order to protect its leg. Later the wound will heal. How different from what happens if you cut the leg of a table. The whole table will neither come to the defense of its leg, nor will it help it to heal. In the more intelligent living creatures, there is also a very great deal of interdependence between an animal's past experience and its present behavior; that is to say, it learns, its behavior changes with time depending on its experiences. Such enormous interdependence must be a monstrously rare occurrence in a universe, unless it has evolved gradually.

In summary, the case is the whole versus the sum of its parts. If both are equally complex, the parts are independent (do not interact). If the whole is very much simpler than the sum of its parts, we have the interdependence that characterizes a living being.(8) Note finally that we have introduced something new into the study of the size of programs (= complexity). Before we compared the sizes of programs that calculate different things. Now we are interested in comparing the sizes of programs that calculate the same thing in different ways. That is to say, the method by which a calculation is done is now of importance to us; in the previous section it was not.

Section 5. Numerical examples.--

In this paper, unfortunately, we can only suggest a possible point of departure for a mathematical definition of life. A great amount of work must be done; it is not even clear what is the formal mathematical conterpart to the informal definition of the previous section. A possibilty is sketched here.

Consider a computer C1 which accepts programs P which are binary sequences consisting of a number of subsequences B, C, P(1),... P(k), A.

B, the leftmost subsequence, is a program for breaking the remainder of P into C, P(1),... P(k), and A. B is self-describing; it starts with a binary sequence which results from writing the length of B in base-two notation, doubling each of its bits, and then placing a pair of unequal bits at the right end. Also, B is not allowed to actually see whether any of the remaining bits of P are 0's or 1's, only to separate them into groups.(9)

C is the description of a computer C2. For example, C2 could be one of Solomonoff's optimal binary computers, or a computer which emits the program without processing it.

P(1),... P(k) are programs which are processed by k different copies of the computer C2. R(1),... R(k) are the resulting outputs. These outputs would be regions of space-time, a space-time which, like von Neumann's, has been cut up into little cubes with a finite number of states.

A is a program for adding together R(1),... R(k) to produce R, a single region of space-time. A merely juxtaposes the intermediate results R(1),... R(k) (perhaps with some overlapping); it is not allowed to change any of the intermediate results. In the examples below, we shall only compute regions R which are one-dimensional strings of 0's and 1's, so that A need only indicate that R is the concatenation of R(1) to R(k), in that order.

R is the output of the computer C1 produced by processing the program P.

We now define a family of complexity measures C(d,R), the complexity of a region R of space-time when it is viewed as the sum of independent regions of diameter not greater than d. C(d,R) is the length of the shortest program P which makes the computer C1 output R, among all those P such that the intermediate results R(1) to R(k) are all less than or equal to d in diameter. C(d,R) where d equals the diameter of R is to within a bounded difference just the usual Solomonoff complexity measure. But as d decreases, we may be forced to forget any patterns in R that are more than d in diameter, and the complexity C(d,R) increases.

We present below a table with four examples. In each of the four cases, R is a 1-dimensional region, a binary sequence of length n. R1 is a random binary sequence of length n ('gas'). R2 consists of n repetitions of 1 ('crystal'). The left half of R3 is a random binary sequence of length n/2. The right half of R3 is produced by rotating the left half about R3's midpoint ('bilateral symmetry'). R4 consists of two identical copies of a random binary sequence of length n/2 ('twins').

C(d,R) = approx.	R = R1 'gas'	R = R2 'crystal'	R = R3 'bilateral symmetry'	R = R4 'twins'
d = n	n	log2 n note 1	n/2	n/2
d = n/k (k>1 fixed, n large)	n	k log2 n notes 1,2	n-(n/2k) note 2	n note 2
d = 1	n	n	n	n

Note 1. This supposes that n is represented in base-two notation by a random binary sequence. These values are too high in those rare cases where this is not true.
Note 2. These are conjectured values. We can only show that C(d,R) is approximately less than or equal to these values.

Footnotes

1. See in particular his fifth lecture delivered at the University of Illinois in December of 1949, 'Re-evaluation of the problem of complicated automata- Problems of hierarchy and evolution', and his unfinished The Theory of Automata: Construction, Reproduction, Homogeneity. Both are published posthumously in von Neumann (1966).
2. See pages 76-77 of von Neumann (1966).
3. See page 78 of von Neumann (1966).
4. See especially pages 78-80 of von Neumann (1966).
5. The earliest generally available appearance in print of Solomonoff's ideas of which we are aware is Minsky's summary of them on pages 41-43 of Minsky (1962). A more recent reference is Solomonoff (1964).
6. Solomonoff actually proposes weighing together all the theories into the prediction, giving the simplest theories the largest weight.
7. Solomonoff's approach to the size of programs has been extended in Chaitin (1969b) to the speed of programs.
8. The whole cannot be more complex than the sum of its parts, because one of the ways of looking at it is as the sum of its parts, and this bounds its complexity.
9. The awkwardness of this part of the definition is apparently its chief defect.

Bibliography

Arbib, M. A. (1966). 'Simple self-reproducing universal automata,' Information and Control.
Arbib, M. A. (1967). 'Automata theory and development: Part 1,' Journal of Theoretical Biology.
Arbib, M. A. 'Self-reproducing automata--some implications for theoretical biology'.

Biological Sciences Curriculum Study. (1968). Biological Science: Molecules to Man, Houghton Mifflin Co.
Chaitin, G. J. (1966). 'On the length of programs for computing finite binary sequences,' Journal of the Association for Computing Machinery.
Chaitin, G. J. (1969a). 'On the length of programs for computing finite binary sequences: Statistical considerations,' ibid.
Chaitin, G. J. (1969b). 'On the simplicity and speed of programs for computing infinite sets of natural numbers,' ibid.
Chaitin, G. J. (1970). 'On the difficulty of computations,' IEEE Transactions on Information Theory.
Codd, E. F. (1968). Cellular Automata. Academic Press.
Kolmogorov, A. N. (1965). 'Three approaches to the definition of the concept 'amount of information' ', Problemy Peredachi Informatsii.
Kolmogorov, A. N. (1968). 'Logical basis for information theory and probability theory,' IEEE Transactions on Information Theory.
Loveland, D. W. 'A variant of the Kolmogorov concept of complexity,' report 69-4, Math. Dept., Carnegie-Mellon University.
Loveland, D. W. (1969). 'On minimal program complexity measures,' Conference Record of the ACM Symposium on Theory of Computing, May 1969.
Martin-Lof, P. (1966). 'The definition of random sequences,' Information and Control.
Minsky, M. L. (1962). 'Problems of formulation for artificial intelligence,' Mathematical Problems in the Biological Sciences, American Math. Society.
Moore, E. F. (1962). 'Machine models of self-reproduction,' ibid.
Von Neumann, J. (1966). Theory of Self-Reproducing Automata. (Edited by A. W. Burks) University of Illinois Press.
Solomonoff, R. J. (1964). 'A formal theory of inductive inference,' Information and Control.
Willis, D. G. (1969). 'Computational complexity and probability constructions,' Stanford University.

TITLE: STOCHASTIC SOURCES FOR CONTEXT-FREE LANGUAGES

AUTHOR: Sandra E. Hutchins
Department of Applied Physics and Information Science
University of California, San Diego
La Jolla, California 92037

ABSTRACT:

The concept of a stochastic source of a context-free language is presented and one possible souce model based on a probabilistic grammar is defined. It is shown that this source is equivalent to an infinite Markov chain which can be defined recursively. Various statistics of the Markov chain are derived.

An analogue of the Greibach Normal Form theorem is presented, which then allows for the construction of a Markov source with a single terminal symbol per transition. The class of probability distributions generable over context-free languages by such sources is examined.